Empirical Bayes Confidence Estimation

BU-1062-M December 1989

Edward I. George[1]

University of Chicago

George Casella[2]

Cornell University

Key Words and Phrases: Conditional confidence, loss estimation, shrinkage estimation, Stein estimator

AMS 1980 Subject Classification: 62C99, 62F25

[1] Research performed while Professor George was visiting the Mathematical Sciences Institute at Cornell University.

[2] Research supported by National Science Foundation Grant No. DMS89-0039

This paper is also Technical Report No. 81, Graduate School of Business, University of Chicago.

Summary

For the problem of estimating a multivariate normal mean, it is known that confidence sets recentered at shrinkage estimators can offer strictly larger coverage probability than the usual confidence set. Unfortunately, the conventional frequentist report of a constant confidence coefficient (infimum of the coverage probabilities) fails to communicate the gain of these improved confidence sets. Through an empirical Bayes argument we introduce a confidence estimator for the recentered confidence set. The confidence report obtained by this estimator is strictly larger than the conventional infimum report. Furthermore, this confidence estimator is shown to dominate the infimum report according to an appropriate risk criterion. By using this new confidence report, the improved confidence region provides an informative frequentist measure of precision for the shrinkage estimator about which it has been recentered.

1. INTRODUCTION.

A most important companion problem for the point estimation of a multivariate normal mean is to provide associated confidence regions. More precisely, based on the observation of a p–dimensional multivariate normal vector

$$X \sim N_p(\theta, I), \tag{1.1}$$

the classical confidence set for θ is of the form

$$C_0(X) = \{\theta : |\theta - X| \le c\}. \tag{1.2a}$$

For all θ, the coverage probability of this set estimator is constant,

$$P_\theta(\theta \in C_0(X)) \equiv P(\chi^2_p \le c^2) = 1-\alpha, \tag{1.2b}$$

and the practitioner reports that $C_0(X)$ contains θ with confidence $1-\alpha$. Hwang and Casella (1982,1984) showed that the set $C_0(X)$ can be improved by recentering at a positive part Stein estimator

$$\delta_a(X) = u_a(|X|)X, \quad \text{where} \quad u_a(r) = \max\{[1-(a/r^2)],0\} \tag{1.3}$$

to obtain

$$C_a(X) = \{\theta : |\theta - \delta_a(X)| \le c\}. \tag{1.4}$$

This set dominates $C_0(X)$ in the sense that both sets are of equal volume but for a in a certain range, $C_a(X)$ has uniformly higher coverage probability for all θ, that is,

$$P_\theta(\theta \in C_a(X)) > P_\theta(\theta \in C_0(X)) \equiv 1-\alpha. \tag{1.5}$$

For this situation the statistician is better off reporting $C_a(X)$ than $C_0(X)$ as a $1-\alpha$ confidence region for θ. Table 1 and Figure 1 display values of $P_\theta(\theta \in C_{p-2}(X))$ for $1-\alpha =$ .90 and a variety of p and $|\theta|$ so the reader can appreciate the potential improvement offered by $C_a(X)$. Note that the improvement increases as θ gets closer to the shrinkage target 0 where $\delta_a(X)$ provides most shrinkage.

[TABLE 1 AND FIGURE 1 ABOUT HERE]

The conventional frequentist confidence report of the infimum of the coverage probabilities is woefully inadequate for C_a. Since

$$\inf_{\theta} P_{\theta}(\theta \in C_a(X)) = 1-\alpha,$$

both C_a and C_0 have the same infimum confidence report. As is clear from Table 1 and Figure 1, reporting confidence $(1-\alpha)$ in C_a is not only misleading, but also fails to communicate to the user the potential improvement offered by C_a. It is the purpose of this paper to develop an alternative (post–experimental) report to $1-\alpha$ which better reflects the coverage of $C_a(X)$. By using this new confidence report, $C_a(X)$ can provide an informative frequentist measure of precision for the Stein estimator $\delta_a(X)$.

In order to provide a framework for the selection of an alternative report, we treat the problem of reporting confidence as an additional estimation problem. More formally, we consider a confidence procedure as a pair $<C(X),\gamma(X)>$ where $C(X)$ is a set estimator and $\gamma(X)$ is a quoted confidence estimator, the confidence that θ is in the set $C(X)$. Such an approach is natural in both the Bayesian and conditional frequentist setting (see Kiefer (1977) and Berger and Wolpert (1984)). The estimator $\gamma(X)$ can be evaluated by a frequentist approach using the **communication risk**

$$R(\theta,<C(X),\gamma(X)>) = E_{\theta}\Big[I[\theta \in C(X)] - \gamma(X)\Big]^2. \tag{1.6}$$

This criterion evaluates $\gamma(X)$ as an estimate of $I[\theta \in C(X)]$, ($I[\cdot]$ is the indicator function), under expected squared error loss. As usual, a confidence function $\gamma_1(X)$ dominates $\gamma_2(X)$ for the set estimator $C(X)$ if

$$R(\theta,<C(X),\gamma_1(X)>) \leq R(\theta,<C(X),\gamma_2(X)>) \tag{1.7}$$

for all θ with strict inequality for some θ. We note that squared error loss is a proper scoring rule here. In the Bayesian setting, minimizing the posterior risk with this loss would yield the posterior probability of $C(X)$ as the optimal γ.

Lu and Berger (1989) addressed this problem and demonstrated that for the recentered set estimator $C_a(X)$, there exists a confidence function of the form

$$\gamma_{LB}(X) = 1-\alpha + \frac{bp}{dp+|X|^2}\alpha \tag{1.8}$$

which, for some positive b and d, uniformly dominates $1-\alpha$ in terms of communication risk.

This confidence function has the intuitively desirable properties that $\gamma_{LB}(X) > 1-\alpha$ and $\gamma_{LB}(X)$ increases with the amount of shrinkage provided by $\delta_a(X)$. Lu and Berger also show that γ_{LB} has the property of ***frequentist validity***, that is,

$$E_\theta \gamma_{LB}(X) \le P_\theta(\theta \in C_a(X)) \quad \text{for all } \theta, \tag{1.9}$$

a property of long run conservative behavior.

Through an empirical Bayes argument, we introduce another confidence function for $C_a(X)$ with many desirable properties. This function is given by

$$\gamma_{b,d}(X) = P[\chi_p^2 \le c^2/u_{b,d}(|X|)] \tag{1.10a}$$

where for some constants $b > 0$ and $d \in [0,1)$,

$$u_{b,d}(r) = \max\{[1-(b/r^2)],d\} \tag{1.10b}$$

is a truncated version of the shrinkage factor of the James–Stein estimator, which is truncated at d rather than 0. Note that $u_{a,0}(|X|) \equiv u_a(|X|)$ is exactly the shrinkage factor in $\delta_a(X)$. (When $u_b(|X|) = 0$, we define $c^2/u_b(|X|) = \infty$). The confidence function $\gamma_{b,d}(X)$ has similar properties to $\gamma_{LB}(X)$ in (1.8): $\gamma_{b,d}(X) > 1-\alpha$, and $\gamma_{b,d}(X)$ increases with the amount of shrinkage provided by $\delta_a(X)$. In Section 2, the details of the empirical Bayes derivation of $\gamma_{b,d}$ are given. In Section 3, a comparison of $P_\theta(\theta \in C_a(X))$ and $E_\theta\gamma_{b,d}(X)$ for large $|\theta|$, suggests appropriate choice of b. A minimum value of d is also suggested. In Section 4, we show that for some a,b,d, the confidence procedure $\langle C_a(X),\gamma_{b,d}(X)\rangle$ dominates $\langle C_a(X),1-\alpha\rangle$ in terms of communication risk, and that $\gamma_{b,d}(X)$ has the property of frequentist validity. Simulation results in Section 5 show that $\gamma_{b,d}(X)$ compares favorably with γ_{LB} and that the available gains can be substantial. Section 6 contains a discussion of these, and related, results. Necessary technical lemmas have been placed in an Appendix.

2. EMPIRICAL BAYES MOTIVATION

In this section, we motivate the confidence procedure $\langle C_a(X),\gamma_{b,d}(X)\rangle$ defined by (1.4) and (1.10) as an empirical Bayes approximation to a Bayes credible region and its associated posterior coverage probability. The approximated Bayes procedure is obtained by assuming that θ is a realization from a conjugate normal prior

$$\theta \sim N_p(0,\tau^2 I). \tag{2.1}$$

Following the development in Efron and Morris (1973), the posterior distribution for θ under (2.1) is

$$\theta|X \sim N_p((1-B)X,(1-B)I), \quad \text{where } B = 1/(1+\tau^2). \tag{2.2}$$

Based on this distribution the highest posterior density credible interval estimate for θ is

$$C_B(X) = \{\theta : |\theta - (1-B)X| \leq c\} \tag{2.3a}$$

with posterior coverage probability

$$P^{\theta|X}[\theta \in C_B(X)] = P[\chi_p^2 \leq c^2/(1-B)]. \tag{2.3b}$$

Adopting the perspective that the prior (2.1) is correctly specified but that τ^2 is unknown, an empirical Bayes approximation to (2.3) can be obtained by substituting a data based estimate of $(1-B) = \tau^2/(1+\tau^2)$. We consider the class of empirical Bayes estimates of the form

$$(1-\hat{B}) = u_{b,d}(|X|) = \max\{[1 - b/|X|^2],d\}. \tag{2.4}$$

Such estimates may be based on the marginal distribution of X, $X \sim N_p(0,(1+\tau^2)I)$. For example, $u_{p,0}(|X|))$ is obtained by maximum likelihood and $u_{p-2,0}(|X|)$ is obtained by truncation (at 0)) of the unbiased estimate. Morris (1983) considers the choice $u_{b,d}(|X|)$ with $b = (p-2)^2/p$ and $d = 2/p$. Substitution of $u_a(|X|)$ into $C_B(X)$ yields the recentered confidence set $C_a(X)$ in (1.4). Substitution of $u_{b,d}(|X|)$ into $P^{\theta|X}[\theta \in C_B(X)]$ yields the confidence function $\gamma_{b,d}(X)$ in (1.10).

3. MATCHING CONSTANTS

Useful insight into the appropriateness of $\gamma_{b,d}(X)$ for estimating $P_\theta(\theta \in C_a(X))$ is obtained by investigating the behavior when $|\theta|$ is large. Hwang and Casella (1984), extending arguments of Berger (1980), show

$$(3.1)\qquad P_\theta(\theta \in C_a(X)) = 1-\alpha + (a(2(p-2)-a)/p)c^2 f_p(c^2)|\theta|^{-2} + O(|\theta|^{-3})$$

where f_p is the χ_p^2 density. The following result provides the analogous form for the behavior of $\gamma_{b,d}(X)$ for large $|\theta|$.

THEOREM 3.1: For $c^2 \geq p-2$,

$$(3.2)\qquad E_\theta \gamma_{b,d}(X) = 1-\alpha + bc^2 f_p(c^2)|\theta|^{-2} + o(|\theta|^{-3}).$$

PROOF: Let $g_{b,d}(|X|^2) = \gamma_{b,d}(X) - (1-\alpha) = P[c^2 \leq \chi_p^2 \leq c^2/u_{b,d}(|X|)]$. The result now follows from (A.5a) of Lemma A.3. □

(Actually Lemma A.3 shows that $o(|\theta|^{-3})$ can be strengthened to $o(|\theta|^{-4+\epsilon})$ for arbitrary $\epsilon > 0$).

Comparison of (3.1) and (3.2) shows that for large $|\theta|$, $\gamma_{b,d}(X)$ with

$$(3.3)\qquad b = a(2(p-2)-a)/p$$

is the correct estimator (up to second order) for $P_\theta(\theta \in C_a(X))$. In particular, it shows that for the choice $a = p-2$ recommended by Hwang and Casella, one should choose $b = (p-2)^2/p \approx p-2$.

For choosing d, we recommend that $\gamma_{b,d}$ should never be larger than the maximum of $P_\theta(\theta \in C_a(X))$. Otherwise such a report would always overestimate coverage when X was near 0, and could result in negatively biased relevant betting procedures (see Casella, 1987,1988). This maximum coverage probability, which we denote γ_{max}, occurs when $\theta = 0$, and is given by

$$(3.4)\qquad \gamma_{max} = \max_\theta P_\theta(\theta \in C_a(X)) = P_0(0 \in C_a(X))$$

$$= P_0(0 \le |X|^2 \le (c^2+2a+c(c^2+4a)^{1/2})/2),$$

where the last equality in (3.4) follows from Lemma A.8. It then follows from (2.4) and (3.4) that in order for $\gamma_{b,d}(0) = P_0(0 \in C_a)$, we should choose

$$d = d_{min} = \frac{2c^2}{c^2+2a+c(c^2+4a)^{1/2}}. \tag{3.5}$$

The choice $d = d_{min}$ worked well in the simulations presented in Section 5.

4. RISK DOMINATION AND FREQUENTIST VALIDITY

In this section, we focus on the evaluation of $\gamma_{b,d}(X)$ as a confidence report for $C_a(X)$ in terms of its communication risk (1.6). Note that although expressions (3.1) and (3.2) show that for $b = a(2(p-2)-a)/p$, $P_\theta(\theta \in C_a(X))$ and $E_\theta\gamma_{b,d}(X)$ will agree for large $|\theta|$, this is no guarantee that $\gamma_{b,d}(X)$ has good risk properties.

Before stating our main risk evaluation results, we point out that we are only interested in risk assessment for values of a where C_a dominates C_0 in terms of coverage probability. Hwang and Casella (1984) show analytically that the coverage dominance of the recentered region $C_a(X)$ over $C_0(X)$ occurs whenever $a \in (0,a^*]$ for

$$a^* = \min\{a_1,a_2\} \tag{4.1a}$$

where a_1 and a_2 are the unique solutions respectively of

$$\left[\frac{c + (c^2+4a)^{1/2}}{2\sqrt{a}}\right]^{p-2} e^{-c\sqrt{a}/2} = 1 \tag{4.1b}$$

and

$$\left[\frac{(c^2+4a)^{1/2} - c}{2\sqrt{a}}\right] \left[\frac{c + (c^2+4a)^{1/2}}{\sqrt{a}}\right]^{p-1} e^{-c\sqrt{a}} = 1. \tag{4.1c}$$

We should add, however, that numerical calculations of Hwang and Casella strongly suggest that domination holds for values of a larger than p–2, but not as large as 2(p–2), (the bound for point estimation). They recommend the choice a = p–2, mainly due to the optimality of the point estimator for this choice.

Because we are most interested in the comparison of the report $\gamma_{b,d}$ with the conservative report $1-\alpha$, we will express our main results in terms of the risk reduction function,

$$\Delta_\theta(a,b) = R(\theta,<C_a(X),1-\alpha>) - R(\theta,<C_a(X),\gamma_{b,d}(X)>). \tag{4.2}$$

(For notational simplicity, we have suppressed the dependence of Δ_θ on d. We are able to do this because all the results about Δ_θ below are valid for any $d \in [0,1)$). The following result reveals the behavior of $\Delta_\theta(a,b)$ for large $|\theta|$.

THEOREM 4.1: Let f_p be the χ_p^2 density, and Δ_θ given in (4.2). For any $\epsilon > 0$,

$$\Delta_\theta(a,b) = \left[a(2(p-2)-a)/p - b/2\right]2bc^4 f_p^2(c^2)|\theta|^{-4} + O(|\theta|^{-5}). \tag{4.3}$$

PROOF: Let $g_{b,d}(|X|^2) = \gamma_{b,d}(X)-(1-\alpha) = P[c^2 \le \chi_p^2 \le c^2/u_{b,d}(|X|)]$. Then we may express (4.2) as

$$\Delta_\theta(a,b) = = 2E_\theta\left[I[\theta \in C_a(X)] - (1-\alpha) - (1/2)g_{b,d}(|X|^2)\right]g_{b,d}(|X|^2). \tag{4.4}$$

The result now follows from Lemma A.4. □

Theorem 4.1 shows that for $|\theta|$ large, the choice $b = a(2(p-2)-a)/p$ in (3.3), which equates (3.1) and (3.2) up to second order terms, also provides the maximum risk reduction over $1-\alpha$. It also follows from (4.3) that the risk domination of $\gamma_{b,d}(X)$ over $1-\alpha$ will fail when $b \geq 2a(2(p-2)-a)/p$, as H will then be negative for large $|\theta|$.

Of course, it is of more interest to know if $\Delta_\theta(a,b)$ is strictly positive for all θ, for in that case $\gamma_{b,d}$ will uniformly dominate $1-\alpha$ as a confidence report for C_a. The following results show that for any $C_a(X)$ known to dominate C_0 in coverage probability, there exists a confidence function of the form $\gamma_{b,d}(X)$ as in (1.10) which dominates $1-\alpha$ in terms of communication risk.

THEOREM 4.2: For $p \geq 5$ and $a \in (0,a^*]$, there exists $b^* > 0$ such that for all $b \in (0,b^*]$,

$$R(\theta,<C_a(X),\gamma_{b,d}(X)>) \ < \ R(\theta,<C_a(X),1-\alpha>) \text{ for all } \theta.$$

PROOF: Based on the expression (4.2) for the risk difference, we will show that for $a \in (0,a^*]$, there exists $b^* > 0$ such that for all $b \in (0,b^*]$, $\Delta_\theta(a,b) > 0$ for all θ. We make use of the following two facts which are consequences of Theorem 4.1 and Lemma A.6, respectively:

(4.5) for $a \in (0,a^*]$, $b \in (0,b']$ where $b' = 2a(2p-4-a)$, $\lim\limits_{|\theta|\to\infty} |\theta|^4 \Delta_\theta(a,b) > 0$,

(4.6) $(\partial/\partial a)\Delta_\theta(a,b) > 0$ for $a \in (0,a^*]$, b and θ.

Because $\Delta_\theta(0,0) = 0$ for all θ (trivially), it follows from (4.6) that

(4.7) $\Delta_\theta(a,0) > 0$ for all $a \in (0,a^*]$ and all θ.

It follows from (4.5) that there exists K such that

(4.8) $\Delta_\theta(a,b) > 0$ for all $a \in (0,a^*]$, $b \in (0,b']$ and $|\theta| > K$.

Because $\Delta_\theta(a,b)$ is continuous in b, there exists $b'' > 0$ such that

(4.9) $$\Delta_\theta(a,b) > 0 \text{ for all } a \in (0,a^*],\ b \in (0,b'']\text{ and } |\theta| \le K.$$

The desired result now follows from (4.8) and (4.9) by taking $b^* = \min\{b',b''\}$.

□

We should point out that Theorem 4.2 is an existence proof since an explicit bound for b^* is not obtained. This limitation was also a characteristic of the risk dominance result of γ_{LB} in Lu and Berger (1989). One approach to obtaining such a bound for $\gamma_{b,d}$, would be to determine the sign of $(\partial/\partial b)\Delta_\theta(a,b)$ for general b, which coupled with (4.6) would yield the desired result. Unfortunately, this result remains elusive.

The next result shows that for certain choices of a,b,d that $\gamma_{b,d}$ has the property of frequentist validity for C_a.

THEOREM 4.3: For $a \in (0,a^*]$ and $d \in (0,1)$, there exists a b^* such that for all $b \in (0,b^*]$,

$$E_\theta \gamma_{b,d}(X) \le P_\theta(\theta \in C_a(X)) \qquad \text{for all } \theta.$$

PROOF: We begin as in the proof of Theorem 2.1 in Lu and Berger (1989). It follows from (1.5) and (3.1), that for $a \in (0,a^*]$ and $b < p$, there exists $\epsilon > 0$ such that

$$\inf_\theta \{(|\theta|^2+p-b)[P_\theta(\theta \in C_a(X)) - (1-\alpha)]\} \ge \epsilon.$$

Thus,

$$P_\theta(\theta \in C_a(X)) > 1-\alpha + \epsilon/(|\theta|^2+p-b) \text{ for all } \theta.$$

Since $\gamma_{b,d}(X) = 1-\alpha + g_{b,d}(|X|^2)$, it suffices to show that b can chosen so that

$$E_\theta g_{b,d}(|X|^2) \le \epsilon/(|\theta|^2+p-b) \text{ for all } \theta.$$

This is exactly what is proved in Lemma A.7. □

5. NUMERICAL RESULTS

To further investigate the performance of $\gamma_{b,d}$, we carried out simulations to estimate a variety of its characteristics. Based on the considerations described in Section 3 we investigated the case $a = p-2$, $b = (p-2)^2/p$, $d = d_{min}$ (from 3.5). We calculated for a large number of values for $1-\alpha$, p and $|\theta|$, but here we will only report on the case $1-\alpha = .90$ and $p = 5,8,15$. For these cases we computed the actual coverage probability of C_{p-2} for a large number of $|\theta|$ values.

In Table 1 and Figure 1 we compare these coverage probabilities to $E_\theta(\gamma_{b,d}(X))$ and $E_\theta(\gamma_{LB}(X))$. (We use the choice of constants for γ_{LB} recommended by Lu and Berger (1989), but make one further modification which improves the performance of γ_{LB}. We truncate γ_{LB} at γ_{max} of (3.4).) Both the table and the figure show that $\gamma_{b,d}$ and γ_{LB} are close to $P_\theta(\theta \in C_{p-2}(X))$, with $\gamma_{b,d}$ being slightly closer. Table 2 and Figure 2 provide a risk comparison of $\gamma_{b,d}$, γ_{LB} and $1-\alpha$ using the loss (1.6). We see that both $\gamma_{b,d}$ and γ_{LB} provide substantial risk improvement over $1-\alpha$, and are comparable among themselves. These also show that the substantial risk improvement is concentrated where C_{p-2} achieves its greatest improvement in coverage probability, with everything collapsing together as $|\theta|$ becomes large.

[TABLE 2 AND FIGURE 2 ABOUT HERE]

Lastly, Table 3 and Figure 3 show the proportional decrease in communication risk of $\gamma(X)$ over $1-\alpha$, namely

$$\frac{R(\theta,<C_{p-2}(X),1-\alpha>) - R(\theta,<C_{p-2}(X),\gamma(X)>)}{R(\theta,<C_{p-2}(X),1-\alpha>)} \tag{5.1}$$

for $\gamma(X) = \gamma_{b,d}(X)$ and $\gamma(X) = \gamma_{LB}(X)$. The figure clearly shows the advantage that $\gamma_{b,d}$

holds over both γ_{LB} and $1-\alpha$. The proportional decrease of $\gamma_{b,d}$ can be almost 10% better than that of γ_{LB}.

[TABLE 3 AND FIGURE 3 ABOUT HERE]

6. DISCUSSION

Treating the confidence estimation problem in a decision theory framework has enabled us to do away with the conventional infimum report $1-\alpha = \inf_\theta P_\theta(\theta \in C(X))$, and consider more relevant and useful confidence reports. In particular, we have treated the problem as one of estimating "confidence" rather than estimating coverage probabilities. We have equated estimating confidence with estimating the indicator function $I[\theta \in C(X)]$, as in (1.6), using the loss and resulting risk of (1.6). This method of estimating and evaluating confidence is preferable to the estimation of coverage probabilities, and results in acceptable answers, for a number of reasons.

For a fixed set estimator $C(X)$, consider the class of Bayes rules against the loss

$$[I[\theta \in C(X)] - \gamma(X)]^2 . \tag{6.1}$$

For $X \sim f(x|\theta)$ and prior $\pi(\theta)$, the Bayes rule is the posterior probability

$$\gamma^\pi(x) = \frac{\int_C f(x|\theta)\pi(\theta)d\theta}{\int_\Theta f(x|\theta)\pi(\theta)d\theta} . \tag{6.2}$$

So for this estimation problem, that of estimating $I[\theta \in C(X)]$, the Bayes (and thus some admissible) rules against squared error loss are posterior, or post–data, coverage probabilities. Only for this set–up will the Bayes rules be posterior probabilities.

For another contrast between estimating $I[\theta \in C(X)]$ and $P_\theta(\theta \in C(X))$ with an estimator $\gamma(X)$, consider the identity

$$\begin{aligned} E_\theta[I[\theta \in C(X)] - \gamma(X)]^2 = E_\theta[P_\theta(\theta \in C(X)) - \gamma(X)]^2 \\ - 2\mathrm{Cov}[I[\theta \in C(X)], \gamma(X)] \\ + P_\theta(\theta \in C(X))[1 - P_\theta(\theta \in C(X)], \end{aligned} \tag{6.3}$$

where $\mathrm{Cov}[I[\theta \in C(X)], \gamma(X)]$ is the covariance between $I[\theta \in C(X)]$ and $\gamma(X)$. Since the third term in (6.3) is beyond our control, estimation of $I[\theta \in C(X)]$ is composed of estimation of coverage probabilities and an evaluation of covariance. Of course, we want $\mathrm{Cov}[I[\theta \in C(X)], \gamma(X)] \geq 0$, and the loss function will penalize us if this is not the case. This quantity is not taken into account if we only estimate the coverage probability.

The choice of constants for the function $u_{b,d}(r)$ of (1.10b), as detailed in Section 3, attempts to make $\gamma_{b,d}(X)$ mimic the coverage probability. The fact that $\gamma_{b,d}(X)$ is also frequency valid is an interesting added attraction, since the construction of the confidence estimator does not take this property into account. However, we do not view the property of frequency validity as a main desideratum of a confidence estimator, and are really not sure of the overall worth of the property. To deliberately underestimate (on average) the true coverage seems unsound from a strict frequentist view. (This situation should not be confused with that of Stein estimation, where shrinkage towards a point results in biased estimators. Here, frequency validity dictates shrinking away from the parameter.) Although having a confidence estimator that is smaller than the coverage probability is a conservative tactic, it does not lead to optimal behavior against a loss function, and is not a property of admissible rules. It may also lead to non–coherent procedures since, if the true probability is always underestimated, then relevant sets will exist, and conditional inferences will be suspect.

In concluding, this paper has established the merits of using the improved confidence procedure $\langle C_a(X), \gamma_{b,d}(X)\rangle$ over $\langle C_a(X), 1-\alpha\rangle$ and certainly $\langle C_0(X), 1-\alpha\rangle$. This procedure not only offers better coverage properties than $C_0(X)$, it also provides a meaningful report of the improvement. Furthermore, just as $\langle C_0(X), 1-\alpha\rangle$ provides an associated frequentist precision measure for $\delta_0(X) = X$, the procedure $\langle C_a(X), \gamma_{b,d}(X)\rangle$ provides an associated frequentist precision measure for the Stein estimator $\delta_a(X)$. We should mention that it may be possible to improve further on $\gamma_{b,d}$ as a confidence estimator for C_a. However, the empirical Bayes form of $\gamma_{b,d}$ suggests that we may have a first approximation to an admissible rule. Although $\gamma_{b,d}(X)$ is not admissible (it is non analytic), by mimicking a posterior probability we have a post data confidence estimator that performs quite reasonably against frequentist criteria and (we hope) is close to admissible.

REFERENCES

Berger, J.O. (1980). Statistical Decision Theory and Bayesian Analysis; Springer (2nd edition).

Berger, J.O. and R.W. Wolpert (1984). The Likelihood Principle, Second Edition. IMS Monograph Series, Institute of Mathematical Statistics, Hayward, CA.

Brown, L.D. and J.T. Hwang (1989) Estimated confidence approach and the validity admissibility criterion. Technical Report, Statistics Center, Cornell University.

Casella, G. (1987). Conditionally acceptable recentered set estimators. Ann. Stat. 15 1363–1371.

Casella, G. (1988). Conditionally acceptable frequentist solutions. In Statistical Decision Theory and Related Topics IV, Volume I (SS. Gupta and J.O. Berger, eds.) 73–117.

Hwang, J.T. and G. Casella (1982). Minimax confidence sets for the mean of a multivariate normal distribution. Ann. Stat. 10 868–881.

Hwang and Casella (1984). Improved set estimators for a multivariate normal mean. Stat. Dec. Supplement Issue 1 3–16.

Kiefer,J.C. (1977). Conditional confidence statements and confidence estimators (with discussion). J. Amer. Statist. Assoc. 72 789–808.

Lu, K. and J.O. Berger (1989). Estimated confidence procedures for multivariate normal means. J. Stat. Plann. Inf. 23 1–19.

Robert, C. and G. Casella (1989). Improved confidence estimators for the usual multivariate normal confidence set. Biometrics Unit Technical Report BU–1041–M, Cornell University, Ithaca, NY.

APPENDIX: THE LEMMAS

These technical lemmas form the basis of the results of this paper, and have been placed in this appendix to improve the readability of the main text.

LEMMA A.1: For X, $\theta \in \mathbb{R}^p$,

$$|X|^{-2n} = |\theta|^{-2n} - 2n(X-\theta)'\theta|\theta|^{-2(n+1)} - n(X-\theta)'(X-\theta)|\theta^*|^{-2(n+1)}$$

$$+ 2n(n+1)(X-\theta)'\theta^*\theta^{*\prime}(X-\theta)|\theta^*|^{-2(n+2)}$$

where for some $t \in [0,1]$, $\theta^* = t\theta + (1-t)X$.

PROOF: Straightforward application of Taylor's Theorem. □

LEMMA A.2: Let $A_\theta = \{X : |X - \theta| \le \theta^\epsilon\}$. Then for any $\epsilon > 0$,

(A.1a) $$E_\theta(|X|^2-b)^{-1}I[X \in A_\theta] = |\theta|^{-2} + o(|\theta|^{-4+\epsilon}).$$

(A.1b) $$E_\theta(|X|^2-b)^{-2}I[X \in A_\theta] = |\theta|^{-4} + o(|\theta|^{-8+\epsilon}).$$

PROOF: Let $|\theta|$ be large enough so that $|X|^2 > b$ on A_θ. Then on A_θ,

(A.2) $$(|X|^2-b)^{-1} = |X|^{-2}/(1-b|X|^{-2}) = \sum_{n=1}^{\infty} b^{n-1}|X|^{-2n}$$

so that

$$E_\theta(|X|^2-b)^{-1}I[X \in A_\theta] = E_\theta\left[\sum_{n=1}^{\infty} b^{n-1}|X|^{-2n}\right]I[X \in A_\theta]$$

and

$$E_\theta(|X|^2-b)^{-2}I[X \in A_\theta] = E_\theta\left[\sum_{n=1}^{\infty} b^{n-1}|X|^{-2n}\right]^2 I[X \in A_\theta].$$

Thus it suffices to show that for $n \geq 1$,

$$\text{(A.3)} \qquad E_\theta |X|^{-2n} I[X \in A_\theta] = |\theta|^{-2n} + o(|\theta|^{-4n+\epsilon}).$$

From Lemma A.1, we have the Taylor expansion of $|X|^{-2n}$ about $|\theta|^{-2n}$,

$$\text{(A.4)} \qquad |X|^{-2n} = |\theta|^{-2n} + h_1(X,\theta) + h_2(X,\theta)$$

$$h_1(X,\theta) = -2n(X-\theta)'\theta|\theta|^{-2(n+1)}$$
$$h_2(X,\theta) = -n(X-\theta)'(X-\theta)|\theta^*|^{-2(n+1)} + 2n(n+1)(X-\theta)'\theta^*\theta^{*\prime}(X-\theta)|\theta^*|^{-2(n+2)}$$

where for some $t \in [0,1]$, $\theta^* = t\theta + (1-t)X$.

The expectation of the first term on the right of (A.4) is trivially $|\theta|^{-2n}$. Because the distribution of X is symmetric around θ over A_θ, $E_\theta h_1(X,\theta) I[X \in A_\theta] = 0$.

To deal with h_2, note that on A_θ, $|X-\theta| = O(|\theta|^\epsilon)$ and $\theta^* = O(|\theta|)$, so that everywhere on A_θ, $h_2(X,\theta) = O(|\theta|^{-4n+\epsilon})$. Thus,

$$E_\theta h_2(X,\theta) I[X \in A_\theta] = O(|\theta|^{-4n+\epsilon}).$$

Because ϵ is arbitrary we may replace $O(|\theta|^{-4n+\epsilon})$ by $o(|\theta|^{-4n+\epsilon})$. This shows (A.3).

□

LEMMA A.3: Let $g_{b,d}(|X|^2) = P[c^2 \leq \chi_p^2 \leq c^2/u_{b,d}(|X|)]$, and let f_p be the χ_p^2 density. For $c^2 \geq p-2$ and any $\epsilon > 0$, (A.5a) $$E_\theta g_{b,d}(|X|^2) = bc^2 f_p(c^2)|\theta|^{-2} + o(|\theta|^{-4+\epsilon})$$

(A.5b) $$E_\theta g^2_{b,d}(|X|^2) = b^2 c^4 f^2_p(c^2)|\theta|^{-4} + O(|\theta|^{-5})$$

PROOF: Let $A_\theta = \{X : |X - \theta| \leq \theta^\epsilon\}$. Since $g_{b,d} \leq 1$, it follows immediately that

(A.6a) $$E_\theta g_{b,d}(|X|^2)I[X \epsilon A^c_\theta] = O(e^{-\theta^{2\epsilon}/2})$$

(A.6b) $$E_\theta g^2_{b,d}(|X|^2)I[X \epsilon A^c_\theta] = O(e^{-\theta^{2\epsilon}/2})$$

Thus it suffices to show

(A.7a) $$E_\theta g_{b,d}(|X|^2)I[X \epsilon A_\theta] = bc^2 f_p(c^2)|\theta|^{-2} + o(|\theta|^{-4+\epsilon})$$

(A.7b) $$E_\theta g^2_{b,d}(|X|^2)I[X \epsilon A_\theta] = b^2 c^4 f^4_p(c^2)|\theta|^{-4} + O(|\theta|^{-5})$$

Let $|\theta|$ be large enough so that $u_{b,d}(|X|) = (1 - b/|X|^2)$ on A_θ. Since $f_p(y)$ is decreasing for $y \geq p-2$, we may expand $g_{b,d}$ on A_θ as

(A.8a) $$g_{b,d}(|X|^2) = \int_{c^2}^{c^2/u_{b,d}(|X|)} f_p(y)dy$$

$$= f_p(c^2)\left[(c^2/u_{b,d}(|X|) - c^2\right] + R$$

$$= bc^2 f_p(c^2)(|X|^2 - b)^{-1} + O(|X|^{-4})$$

since for some fixed K,

$$\text{(A.8b)} \qquad |R| \le \left[f_p(c^2) - f_p((c^2/u_{b,d}(|X|))\right]\left[(c^2/u_{b,d}(|X|) - c^2\right]$$

$$\le K(|X|^2 - b)^{-2} = O(|X|^{-4}).$$

(A.7a) then follows from applying (A.1a) to (A.8a), and noting that on A_θ, $|X|^{-4} = O(|\theta|^{-4})$. Now observe that from (A.8a)

$$\text{(A.9)} \qquad g^2_{b,d}(X) = b^2 c^4 f^2_p(c^2)(|X|^2 - b)^{-2} + O(|X|^{-5})$$

on A_θ. (A.7b) then follows from applying (A.1b) to (A.9), and noting that on A_θ, $|X|^{-5} = O(|\theta|^{-5})$. □

LEMMA A.4: Let $\Delta_\theta(a,b) = 2E_\theta\left[I[\theta \in C_a(X)] - (1-\alpha) - (1/2)g_{b,d}(|X|^2)\right]g_{b,d}(|X|^2)$ where $g_{b,d}(|X|^2) = \gamma_{b,d}(X) - (1-\alpha) = P[c^2 \le \chi^2_p \le c^2/u_{b,d}(|X|)]$. Then

$$\text{(A.10)} \qquad \Delta_\theta(a,b) = \left[a(2p-4-a)/p - b/2\right]2bc^4 f^2_p(c^2)|\theta|^{-4} + O(|\theta|^{-5}).$$

PROOF: It is easy to see that $\Delta_\theta(a,b)$ may be expressed as

$$\text{(A.11)} \qquad \Delta_\theta(a,b) = \Delta_{\theta,1}(a,b) + \Delta_{\theta,2}(b)$$

where

$$\Delta_{\theta,1}(a,b) = 2E_\theta\left[I[\theta \in C_a(X)] - (1-\alpha)\right]g_{b,d}(|X|^2)$$

$$\Delta_{\theta,2}(b) = -E_\theta g^2_{b,d}(|X|^2)$$

Note that when $I[\theta \in C_a(X)] = 1$, $|X-\theta|$ is bounded by a constant in which case $(|X|^2-b)^{-1} = |\theta|^{-2} + O(|\theta|^{-3})$, which follows from (A.2) and (A.3), and $|X|^{-4} = O(|\theta|^{-4})$. Using (A.8a), we have

$$\text{(A.12)} \qquad g_{b,d}(|X|^2) = bc^2 f_p(c^2)(|X|^2-b)^{-1} + O(|X|^{-4})$$

$$= bc^2 f_p(c^2)|\theta|^{-2} + O(|\theta|^{-3})$$

when $I[\theta \in C_a(X)] = 1$. Thus,

$$\text{(A.13)} \qquad \Delta_{\theta,1}(a,b) = 2E_\theta\Big[I[\theta \in C_a(X)] - (1-\alpha)\Big]\Big[bc^2 f_p(c^2)|\theta|^{-2} + O(|\theta|^{-3})\Big]$$

$$= (2a(2p-4-a)/p)bc^4 f_p^2(c^2)|\theta|^{-4} + O(|\theta|^{-5})$$

where the second equality used the fact that

$$E_\theta\Big[I[\theta \in C_a(X)] - (1-\alpha)\Big] = a(2(p-2)-a)c^2 f_p(c^2)p^{-1}|\theta|^{-2} + O(|\theta|^{-3})$$

which follows from (3.1), obtained by Hwang and Casella (1984). Note that from (A.5b)

$$\text{(A.14)} \qquad \Delta_{\theta,2}(b) = b^2c^4 f_p^2(c^2)|\theta|^{-4} + O(|\theta|^{-5}).$$

Inserting (A.13) and (A.14) into (A.11) yields (A.10). □

LEMMA A.5: Let $h(r^2) = r^2 g_{b,d}(r^2)$ where $g_{b,d}(r^2) = P[c^2 \le \chi_p^2 \le c^2/u_{b,d}(r)]$. For $p \ge 3$ and $c^2 \ge p+2$, $h(r^2)$ is increasing in r^2.

PROOF: For simplicity, let $t = r^2$. For $t \in [0, b/(1-d)]$, $h(t)$ is clearly increasing because $g_{b,d}(t)$ is constant over this range. For $t > b/(1-d)$, it suffices to show that $h''(t) < 0$. This will force the impossibility of $h'(t) < 0$ for any $t > b/(1-d)$, since h would then become negative for large t contradicting the obvious fact that $h(t) \geq 0$ for all t. Straightforward differentiation with respect to t yields

$$\text{(A.15)} \qquad h'(t) = tg'_{b,d}(t) + g_{b,d}(t) = \frac{-tc^2 b}{(t-b)^2} f_p(c^2/u) + g_{b,d}(t)$$

where f_p is the χ^2_p density. Differentiating again yields

$$\text{(A.16)} \qquad h''(t) = \frac{c^2 b}{(t-b)^4}\left[tc^2 bf'_p(c^2/u) + (t^2-b^2)f_p(c^2/u) - (t-b)^2 f_p(c^2/u)\right].$$

Making use of the identity $f_p(w) = (f_{p-2}(w) - f_p(w))/2$ and the expression $f_p(w) = \left[\Gamma(p/2)\, 2^{p/2}\right]^{-1} w^{(p-2)/2}\, e^{-w/2}$, the expression in brackets in (A.16) is

$$\text{(A.17)} \qquad \left[\;\right] = (tc^2 b/2)((f_{p-2}(c^2/u) - f_p(c^2/u)) + 2b(t-b)f_p(c^2/u)$$

$$= \frac{(c^2/u)^{(p-4)/2} e^{-c^2/2u}}{\Gamma((p-2)/2)2^{(p-2)/2}}\left[\frac{tc^2 b}{2}(1 - \frac{c^2/u}{p-2}) + 2b(t-b)\frac{c^2/u}{p-2}\right].$$

Recalling that $u = (t-b)/t$, the term in brackets in (A.17) is

$$\text{(A.18)} \qquad \left[\;\right] = \frac{tc^2 b}{2}(1 - \frac{c^2}{p-2}\frac{t}{t-b}) + 2b(t-b)\frac{c^2}{p-2}\frac{t}{t-b} = \frac{tc^2 b}{2(p-2)}\left[p + 2 - \frac{tc^2}{t-b}\right].$$

Now (A.18) will be negative if $t > b$ and $c^2 > p+2$. But this is precisely when $h''(b)$ is negative. □

LEMMA A.6: For a^* as in (4.1) and Δ_θ as in (4.2), $(\partial/\partial a)\Delta_\theta(a,b) > 0$ for all $a \in (0,a^*]$, b and θ.

PROOF: We proceed as in Hwang and Casella (1984) where a more detailed presentation of some of the steps below can be found. We begin with the spherical transformation where $r = |X|$ and β is the angle between X and θ. In terms of these coordinates, the term involving a in Δ_θ as given in (4.3) may be reexpressed as

$$\text{(A.19)} \quad 2E_\theta I[\theta \in C_a(X)]g_{b,d}(|X|^2) = 2K \int_0^{\beta_0}\int_{r_-}^{r_+} r^{p-1}\sin^{p-2}\beta\, f^*(r,\beta)g_{b,d}(r^2)dr d\beta$$

where $K = 2\prod_{i=0}^{p-3}\int_0^\pi \sin^i t dt$,

$$\text{(A.20)} \qquad f^*(r,\beta) = (2\pi)^{-p/2}\exp\{-(r^2-2r|\theta|\cos\beta+|\theta|^2)/2\},$$

and β_0, r_+ and r_- are defined as follows. Defining first r_+^0 and r_-^0 by

$$\text{(A.21)} \qquad r_\pm^0 = |\theta|\cos\beta \pm (c^2 - |\theta|^2\sin^2\beta)^{1/2},$$

for $|\theta| > c$, $\beta_0 = \arcsin(c/|\theta|)$ and

$$\text{(A.22)} \qquad r_\pm \equiv r_\pm(\theta,a,\beta) = \frac{1}{2}\left[r_\pm^0 + [(r_\pm^0)^2 + 4a]^{1/2}\right],$$

whereas for $|\theta| \le c$, $\beta_0 = \pi$, $r\_ = 0$, and

$$r_+ \equiv r_+(\theta,a,\beta) = \frac{1}{2}\left[r_+^0 + [(r_+^0)^2 + 4a]^{1/2}\right].$$

For $|\theta| > c$, $(\partial/\partial a)[r_\pm(\theta,a,\beta)] = 1/\{r_\pm[1+(a/r_\pm^2)]\}$. For $|\theta| \le c$, $(\partial/\partial a)[r_+(\theta,a,\beta)] = 1/\{r_+[1+(a/r_+^2)]\}$ and $(\partial/\partial a)[r\_(\theta,a,\beta)] \equiv 0$. Making use of these expressions, straightforward differentiation yields

$$\text{(A.23)}\quad (\partial/\partial a)\Delta_\theta(a,b) = 2K\int_0^{\beta_0} \sin^{p-2}\beta\left[\frac{r_+^p f^*(r_+,\beta)\, g_{b,d}(r_+^2)}{(a+r_+^2)} - \frac{r_-^p f^*(r_-,\beta) g_{b,d}(r_-^2)}{(a+r_-^2)}\right] d\beta.$$

For the case $|\theta| \le c$, the integrand in (A.23) is positive so that the claim is true. For the case $|\theta| > c$, it is sufficient to prove

$$\text{(A.24)}\qquad \frac{r_+^p f^*(r_+,\beta)(a+r_-^2) g_{b,d}(r_+^2)}{r_-^p f^*(r_-,\beta)(a+r_+^2) g_{b,d}(r_-^2)} > 1, \qquad \text{a.e.}$$

By Lemma A.5, $r^2 g_{b,d}(r^2)$ is increasing in r^2 so that $r_+^2 g_{b,d}(r_+^2)/r_-^2 g_{b,d}(r_-^2) > 1$. Thus it suffices to show that for $a \in (0,a^*]$,

$$\text{(A.25)}\qquad \frac{r_+^{p-2} f^*(r_+,\beta)(a+r_-^2)}{r_-^{p-2} f^*(r_-,\beta)(a+r_+^2)} > 1, \qquad \text{a.e.}$$

But (A.25) is exactly what is shown for $p \ge 5$ in Theorem 2.2 of Hwang and Casella (1984) where the expressions for a^* in (4.1) are given. (Hwang and Casella use p instead of p–2 in (A.25) and show the result for $p \ge 3$). □

A crucial step in proving Lemma A.6 consisted of using the fact that $r^2 g_{b,d}(r^2)$ is increasing in r^2, the consequence of Lemma A.5. This reduced the proof to showing that (A.25) held, which was true for $p \geq 5$. It is interesting to note that $r g_{b,d}(r^2)$ is not increasing in r^2 so that our proof would not work for $p \leq 4$. Indeed, the results of Hwang and Brown (1989) suggest that Theorem 4.2 cannot hold for $p \leq 4$.

LEMMA A.7. For any $\epsilon > 0$ and $d \in (0,1)$, there exists a b^* such that for all $b \in (0,b^*]$,

$$E_\theta g_{b,d}(|X|^2) \leq \epsilon/(|\theta|^2+p-b) \text{ for all } \theta. \tag{A.26}$$

PROOF: Following an argument similar to (A.8), we have that for fixed $K_1,K_2 > 0$,

$$g_{b,d}(|X|^2) \leq K_1\Big[1/u_{b,d}(|X|)-1\Big] + K_2\Big[1/u_{b,d}(|X|)-1\Big]^2.$$

Letting $s = bd/(1-d)$ (or equivalently $d = s/(b+s)$),

$$\Big[1/u_{b,d}(|X|)-1\Big] = (b/s)I[|X|^2 \leq b+s] + b(|X|^2-b)^{-1}I[|X|^2 > b+s].$$

Thus

$$E_\theta g_{b,d}(|X|^2) \leq G_1(b,\theta) + G_2(b,\theta) \tag{A.27a}$$

$$G_1(b,\theta) = [(b/s)K_1 + (b/s)^2K_2]\, P_\theta[|X|^2 \leq b+s] \tag{A.27b}$$

$$G_2(b,\theta) = E_\theta\Big[bK_1(|X|^2-b)^{-1} + b^2K_2(|X|^2-b)^{-2}\Big]I[|X|^2 > b+s]. \tag{A.27c}$$

Using the fact that for fixed b, $P_\theta[|X|^2 \leq b+s]$ is continuous in $|\theta|$ and $P_\theta[|X|^2 \leq b+s] =$

$O(e^{-\theta^2/2})$, it is straightforward to show that we may choose $b_0 > 0$ small enough so that for all $b \in (0,b_0]$,

(A.28) $$G_1(b,\theta) \le \epsilon/2(|\theta|^2+p-b) \le \epsilon/2(|\theta|^2+p-b_0) \text{ for all } \theta.$$

Now note that for all $b \in (0,b_0]$, we have for $n = 1,2$

(A.29) $$E_\theta(|X|^2-b)^{-n}I[|X|^2 > b+s] = E_0(|X+\theta|^2-b)^{-n}I[|X+\theta|^2 > b+s]$$

$$= (|\theta|^2+p-b)^{-n}\left[1 - E_0\left[\frac{|X|^2+2X'\theta-p}{|X+\theta|^2-b}\right]^n I[|X+\theta|^2 > b+s]\right]$$

$$\le (1+\epsilon_n)(|\theta|^2+p-b_0)^{-n}$$

where

(A.30) $$\epsilon_n = E_0\left[2|X|/(\sqrt{b_0+s} - \sqrt{b_0}) + (|X|^2+p)/s\right]^n$$

The last inequality in (A.29) follows by noting that $|X|^2-p+2X'\theta = 2X'(X+\theta) - (|X|^2+p)$ so that on the set $\{X: |X+\theta|^2 > b+s\}$,

(A.31) $$\left|\frac{|X|^2-p+2X'\theta}{|X+\theta|^2-b}\right| \le \left|\frac{2X'(X+\theta)}{|X+\theta|^2-b}\right| + \frac{|X|^2+p}{s}$$

$$\le \left|\frac{2|X||X+\theta|}{(|X+\theta|+\sqrt{b})(|X+\theta|-\sqrt{b})}\right| + \frac{|X|^2+p}{s}$$

$$\le 2|X|/(\sqrt{b_0+s} - \sqrt{b_0}) + (|X|^2+p)/s$$

where we have used the fact that $(\sqrt{b+s}-\sqrt{b}) \geq (\sqrt{b_0+s}-\sqrt{b_0})$ for all $b \in (0,b_0]$. It then follows from (A.31) that

$$\left|E_0\left[\frac{|X|^2+2X'\theta-p}{|X+\theta|^2-b}\right]^n I[|X+\theta|^2 > b+s]\right| \leq \epsilon_n.$$

Now define $b^* = \min\{b_0, \epsilon/4K_1(1+\epsilon_1), \sqrt{\epsilon/4K_2(1+\epsilon_2)}\}$. For b_0 small enough so that $p-b_0 \geq 1$, application of (A.29) to (A.27c) shows that

$$G_2(b,\theta) \leq \epsilon/2(|\theta|^2+p-b_0) \text{ for all } \theta, \tag{A.32}$$

whenever $b \in (0,b^*]$. Hence (A.26) holds for all $b \in (0,b^*]$. □

LEMMA A.8: $P_0(0 \in C_a(X)) = P_0(0 \leq |X|^2 \leq (c^2+2a+c(c^2+4a)^{1/2})/2)$.

PROOF: $P_0(0 \in C_a(X)) = P_0(|\delta_a(X)|^2 \leq c^2)$

$$= P_0((1-a/|X|^2)^2|X|^2 \leq c^2, |X|^2 > a) + P_0(|X|^2 < a)$$

The desired result now follows by noting that

$$\{x: (1-a/x^2)^2x^2 \leq c^2\} = \{x: (c^2+2a-c(c^2+4a)^{1/2})/2 \leq x^2 \leq (c^2+2a+c(c^2+4a)^{1/2})/2\}$$

and

$$(c^2+2a-c(c^2+4a)^{1/2})/2 \leq a \leq (c^2+2a+c(c^2+4a)^{1/2})/2.$$

□

Table 1: For $1\text{-}\alpha = .9$, coverage probabilities of $C_{p\text{-}2}$ and expected values of confidence estimators $\gamma_{b,d}(X)$ and $\gamma_{LB}(X)$. The constants used are a=p-2, b=$(p\text{-}2)^2/p$, and d of (3.5). The constants for $\gamma_{LB}(X)$ are those recommended by Lu and Berger. Both $\gamma_{b,d}(X)$ and $\gamma_{LB}(X)$ were truncated at γ_{max} of (3.4).

	p=5			p=8			p=15		
$\lvert\theta\rvert$	P_θ	$\gamma_{b,d}$	γ_{LB}	P_θ	$\gamma_{b,d}$	γ_{LB}	P_θ	$\gamma_{b,d}$	γ_{LB}
0.0	.988	.988	.988	.998	.998	.998	1.00	1.00	1.00
1.0	.986	.983	.982	.997	.996	.994	1.00	1.00	.999
2.0	.980	.970	.967	.995	.992	.985	1.00	1.00	.996
3.0	.968	.951	.950	.992	.982	.974	1.00	1.00	.992
4.0	.932	.935	.937	.973	.967	.961	1.00	.999	.987
5.0	.921	.924	.927	.956	.953	.950	.999	.997	.981
6.0	.916	.917	.920	.943	.941	.941	.996	.993	.974
7.0	.913	.913	.915	.935	.931	.933	.990	.987	.968
8.0	.911	.910	.912	.928	.925	.927	.983	.979	.961
9.0	.909	.907	.909	.922	.920	.922	.973	.970	.955
10.0	.907	.906	.908	.918	.916	.918	.964	.961	.949
15.0	.903	.902	.903	.909	.906	.908	.931	.930	.926
20.0	.902	.901	.901	.904	.903	.903	.914	.913	.913
25.0	.901	.900	.900	.902	.901	.901	.905	.905	.905
30.0	.900	.900	.900	.900	.900	.900	.900	.900	.900

Figure 1: For $1\text{-}\alpha = .9$, coverage probabilities of C_{p-2} (dotted lines) and expected values of confidence estimators $\gamma_{b,d}(X)$ (solid lines) and $\gamma_{LB}(X)$ (dashed lines). The constants used are a=p-2, $b=(p-2)^2/p$, and d of (3.5). The constants for $\gamma_{LB}(X)$ are those recommended by Lu and Berger. Both $\gamma_{b,d}(X)$ and $\gamma_{LB}(X)$ were truncated at γ_{max} of (3.4).

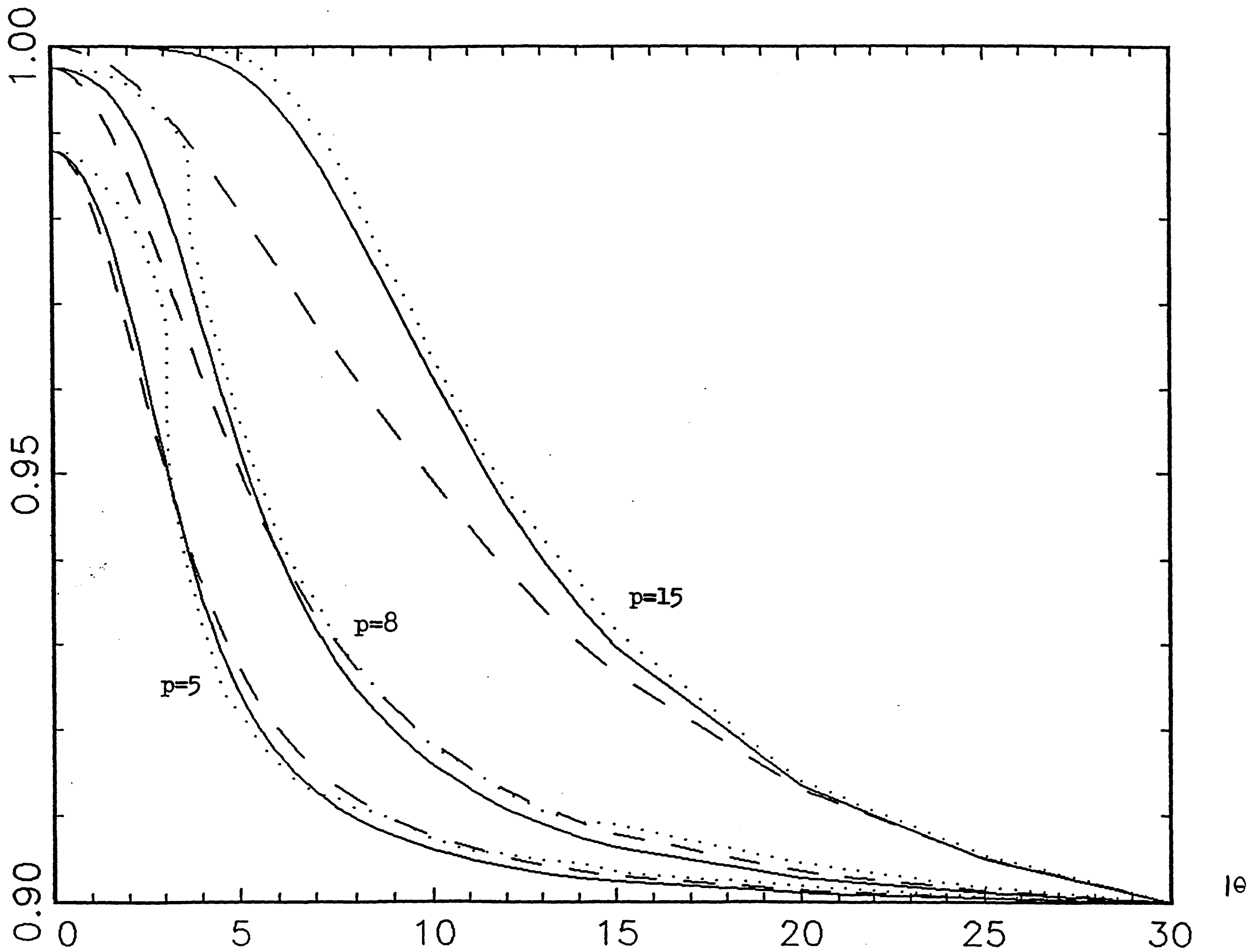

Table 2: For 1-α = .9, risk of the confidence estimators 1-α, $\gamma_{b,d}(X)$ and $\gamma_{LB}(X)$ using squared error loss. The constants used are a=p-2, b=$(p-2)^2/p$, and d of (3.5). The constants for $\gamma_{LB}(X)$ are those recommended by Lu and Berger. Both $\gamma_{b,d}(X)$ and $\gamma_{LB}(X)$ were truncated at γ_{max} of (3.4).

	p=5			p=8			p=15		
$\lvert\theta\rvert$	1-α	$\gamma_{b,d}$	γ_{LB}	1-α	$\gamma_{b,d}$	γ_{LB}	1-α	$\gamma_{b,d}$	γ_{LB}
0.0	.0198	.0113	.0124	.0123	.0027	.0035	.0100	.0000	.0002
1.0	.0215	.0133	.0145	.0126	.0030	.0040	.0100	.0000	.0003
2.0	.0262	.0191	.0204	.0140	.0047	.0059	.0100	.0000	.0004
3.0	.0357	.0308	.0317	.0168	.0083	.0096	.0100	.0000	.0005
4.0	.0643	.0648	.0641	.0323	.0275	.0282	.0101	.0001	.0008
5.0	.0729	.0735	.0730	.0454	.0433	.0431	.0104	.0005	.0015
6.0	.0772	.0775	.0773	.0559	.0551	.0547	.0130	.0038	.0049
7.0	.0800	.0802	.0801	.0623	.0619	.0616	.0176	.0095	.0105
8.0	.0814	.0815	.0814	.0680	.0678	.0676	.0236	.0169	.0177
9.0	.0825	.0826	.0826	.0722	.0720	.0719	.0314	.0263	.0268
10.0	.0843	.0843	.0843	.0758	.0757	.0757	.0390	.0352	.0354
15.0	.0873	.0873	.0873	.0830	.0830	.0830	.0648	.0639	.0639
20.0	.0886	.0886	.0886	.0865	.0864	.0864	.0788	.0786	.0786
25.0	.0896	.0896	.0896	.0887	.0887	.0887	.0857	.0856	.0856
30.0	.0900	.0900	.0900	.0900	.0900	.0900	.0900	.0900	.0900

Figure 2: For 1-α = .9, risk of the confidence estimators 1-α (dotted lines), $\gamma_{b,d}(X)$ (solid lines) and $\gamma_{LB}(X)$ (dashed lines) using squared error loss. The constants used are a=p-2, b=$(p-2)^2/p$, and d of (3.5). The constants for $\gamma_{LB}(X)$ are those recommended by Lu and Berger. Both $\gamma_{b,d}(X)$ and $\gamma_{LB}(X)$ were truncated at γ_{max} of (3.4).

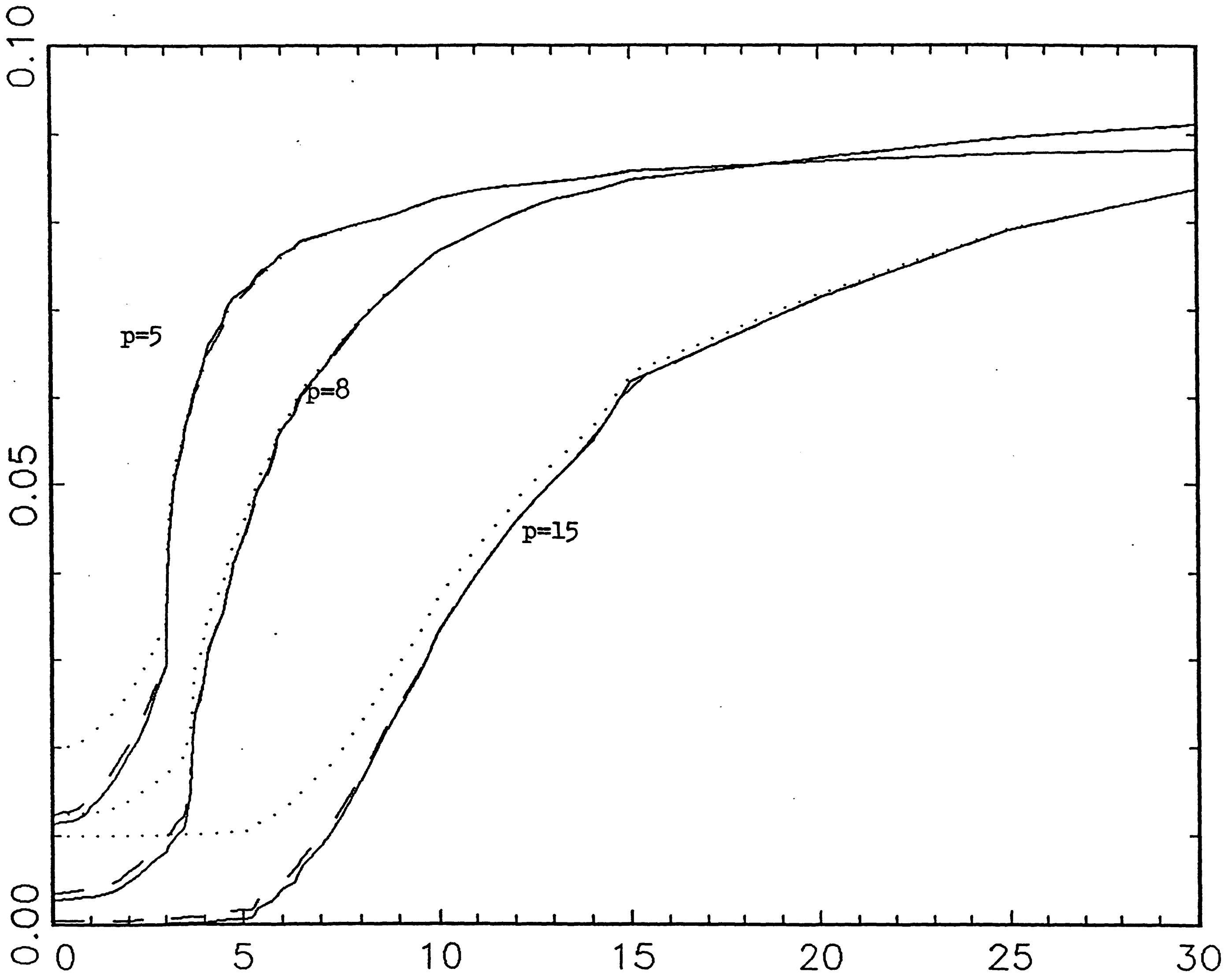

Table 3: For $1\text{-}\alpha = .9$, proportional decrease in risk (5.1) of the confidence estimators $\gamma_{b,d}(X)$ and $\gamma_{LB}(X)$ over $1\text{-}\alpha$. The constants used are a=p-2, $b=(p-2)^2/p$, and d of (3.5). The constants for $\gamma_{LB}(X)$ are those recommended by Lu and Berger. Both $\gamma_{b,d}(X)$ and $\gamma_{LB}(X)$ were truncated at γ_{max} of (3.4).

	p=5		p=8		p=15	
$\lvert\theta\rvert$	$\gamma_{b,d}$	γ_{LB}	$\gamma_{b,d}$	γ_{LB}	$\gamma_{b,d}$	γ_{LB}
0.0	.431	.376	.780	.719	1.00	.971
1.0	.384	.328	.755	.687	1.00	.969
2.0	.273	.224	.660	.580	1.00	.962
3.0	.141	.115	.505	.426	1.00	.949
4.0	-.008	.003	.146	.125	.990	.921
5.0	-.007	-.001	.044	.048	.952	.864
6.0	-.004	-.001	.013	.020	.725	.643
7.0	-.002	-.001	.006	.011	.481	.423
8.0	-.001	-.000	.003	.005	.303	.269
9.0	-.000	.000	.002	.003	.178	.162
10.0	.000	.000	.000	.001	.109	.103
15.0	.000	.000	.000	.000	.018	.019
20.0	.000	.000	.000	.000	.004	.005
25.0	.000	.000	.000	.000	.001	.001
30.0	.000	.000	.000	.000	.000	.000

Figure 3: For 1-α = .9, proportional decrease in risk (5.1) of the confidence estimators $\gamma_{b,d}(X)$ (solid lines) and $\gamma_{LB}(X)$ (dashed lines) over 1-α. The constants used are a=p-2, b=$(p-2)^2/p$, and d of (3.5). The constants for $\gamma_{LB}(X)$ are those recommended by Lu and Berger. Both $\gamma_{b,d}(X)$ and $\gamma_{LB}(X)$ were truncated at γ_{max} of (3.4).